Bitcoin Price & Trend Prediction

Eleftherios Kanavakis

Natural Language Processing for Law and Social Science

2023/06/20

Motivation

- Cryptocurrencies are an emerging financial market that is gaining a lot of traction
- Cryptocurrencies are high volatile assets, thus predicting their behavior is important to protect investors
- Recent advancements in NLP enabled the combination of financial and textual data for better market forecasting
- Various papers claim "crazy" results but can we really beat the markets?

Research Question

- Can we predict both the future trend and price of Bitcoin ?
  - Can results from literature get recreated ?
  - How well could the recent bear market get predicted (March 2022 till today)?
  - Does textual data help with price forecasting?

Related Work

- [Ortu et al](), combined financial & textual data to predict the trend of BTC and ETH using data from 2017 to 2021, achieving 0.83 f1 score
- [Tran et al](), did a survey on the emerging field of predicting digital asset price using NLP
- [Sul et al](), used tweets and financial data to predict stock returns achieving annual economic gains of 11-15%

Financial Data

Bitcoin Historical data were sourced from Bitfinex, a cryptocurrency CEX.

- Hourly candles will be used.
- Dataset ranges from 2018/05 to 2023/06.
- 1859 days and 44635 hours are recorded.
- Lowest BTC price : 3215.2 $
- Highest BTC price : 68958 $

Dataset features (technical):
- Open price
- Close price
- High
- Low
- Date
- Volume BTC
- Volume USD

Data overview

BTC/USD price in time
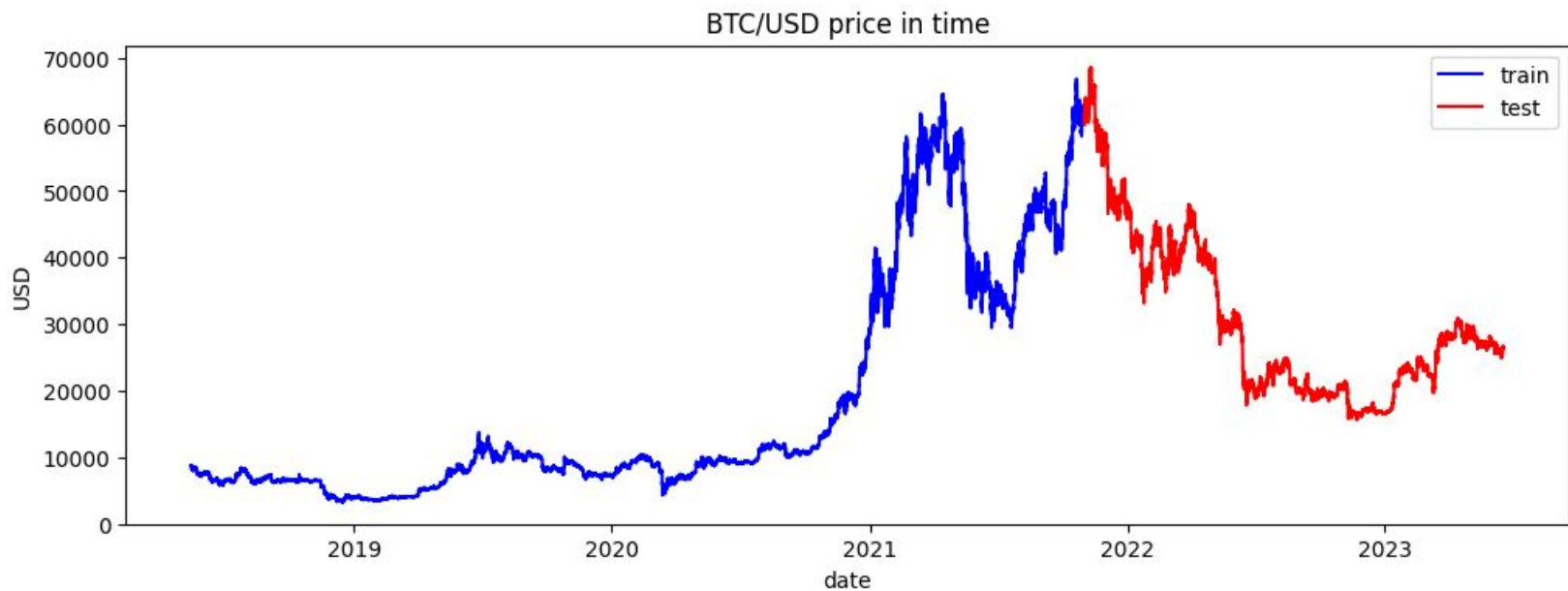
Tasks

Task 1 – BTC trend prediction:
- Binary classification
- Given timestep *t* predict close trend of timestep *t+1*
- If close[t] < close[t+1] **uptrend**
- If close[t] >= close[t+1] **downtrend**

Task 2 – BTC price prediction:
- Regression
- Given timestep *t* predict close price of timestep *t+1*

Methods

Feature engineering :
1. Original features
2. Original features + (year,month,day)
3. Original features + trading indicators (moving averages etc)

Modelling :
1. Xgboost
2. MLP
3. LSTM

Results

Trend prediction (F1):

Xgboost	0.5
MLP	0.49
LSTM	0.57

Trend prediction (√mse):

Xgboost	818.52
MLP	238.98
LSTM	255.76

- Adding year,month,day features did not boost performance
- Trading indicators lead to a 10% performance increase

What's next ?!

- Evaluate the variance of financial models
- Source textual data from Twitter and Reddit regarding BTC
- Perform sentiment analysis using transformers
- Combine financial & textual models and evaluate (**challenge**)

Changes since outline

- Changed from Ethereum to Bitcoin (more data available)
- Downloaded financial data instead of crawling
- Added the task of price forecasting (more challenging)
- Focused more in the bear market (first real cryptocurrency bear market that is supported by macro economic data)

RoadMap

1. Finalize financial models (by June 23rd)
2. Analyze textual data (by July 1st)
3. Perform sentiment analysis (by July 7th)
4. Combine textual & financial models (by July 12th)
5. Hand In rough draft (by July 15th)